

University of Groningen

**Interobserver variation in the classification of thymic lesions including biopsies and resection specimens in an international digital microscopy panel**

Wolf, Janina L.; van Nederveen, Francien; Blaauwgeers, Hans; Marx, Alexander; Nicholson, Andrew G.; Roden, Anja C.; Stroebel, Philipp; Timens, Wim; Weissferdt, Annika; von der Thusen, Jan

*Published in:*  
Histopathology

*DOI:*  
[10.1111/his.14167](https://doi.org/10.1111/his.14167)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Wolf, J. L., van Nederveen, F., Blaauwgeers, H., Marx, A., Nicholson, A. G., Roden, A. C., Stroebel, P., Timens, W., Weissferdt, A., von der Thusen, J., & den Bakker, M. A. (2020). Interobserver variation in the classification of thymic lesions including biopsies and resection specimens in an international digital microscopy panel. *Histopathology*, 77(5), 734-741. <https://doi.org/10.1111/his.14167>

**Copyright**





Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Interobserver variation in the classification of thymic lesions including biopsies and resection specimens in an international digital microscopy panel

Janina L Wolf,<sup>1</sup> Francien van Nederveen,<sup>2</sup> Hans Blaauwgeers,<sup>3</sup> Alexander Marx,<sup>4</sup> Andrew G Nicholson,<sup>5,6</sup>  Anja C Roden,<sup>7</sup>  Philipp Ströbel,<sup>8</sup> Wim Timens,<sup>9</sup> Annika Weissferdt,<sup>10</sup> Jan von der Thüsen<sup>1</sup>  & Michael A den Bakker<sup>1,11</sup> 

<sup>1</sup>Department of Pathology, Erasmus MC, Rotterdam, <sup>2</sup>PAL Laboratory, Dordrecht, <sup>3</sup>Department of Pathology, OLVG Lab BV, Amsterdam, The Netherlands, <sup>4</sup>Universitätsklinikum Mannheim, Heidelberg University, Heidelberg, Germany, <sup>5</sup>National Heart and Lung Institute, Imperial College London, <sup>6</sup>Department of Histopathology, Royal Brompton & Harefield NHS Foundation Trust, London, UK, <sup>7</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, <sup>8</sup>Universitätsklinikum Göttingen, Göttingen, Germany, <sup>9</sup>University Medical Centre Groningen, Groningen, The Netherlands, <sup>10</sup>MD Anderson Cancer Center, Houston, TX, USA, and <sup>11</sup>Maasstad Ziekenhuis, Rotterdam, The Netherlands

Date of submission 18 March 2020

Accepted for publication 31 May 2020

Published online Article Accepted 7 June 2020

Wolf J L, van Nederveen F, Blaauwgeers H, Marx A, Nicholson A G, Roden A C, Ströbel P, Timens W, Weissferdt A, von der Thüsen J & den Bakker M A.

(2020) *Histopathology* 77, 734–741. <https://doi.org/10.1111/his.14167>

## Interobserver variation in the classification of thymic lesions including biopsies and resection specimens in an international digital microscopy panel

**Aims:** Thymic tumours are rare in routine pathology practice. Although the World Health Organization (WHO) classification describes a number of well-defined categories, the classification remains challenging. The aim of this study was to investigate the reproducibility of the WHO classification among a large group of international pathologists with expertise in thymic pathology and by using whole slide imaging to facilitate rapid diagnostic turnover.

**Methods and results:** Three hundred and five tumours, consisting of 90 biopsies and 215 resection specimens, were reviewed with a panel-based virtual microscopy approach by a group of 13 pathologists with expertise in thymic tumours over a period of 6 years. The specimens were classified according to the WHO 2015 classification. The data were subjected to statistical analysis, and interobserver concordance (Fleiss kappa) was calculated. All cases were diagnosed within a time frame of 2 weeks. The

overall level of agreement was substantial ( $\kappa = 0.6762$ ), and differed slightly between resection specimens ( $\kappa = 0.7281$ ) and biopsies ( $\kappa = 0.5955$ ). When analysis was limited to thymomas only, and they were grouped according to the European Society for Medical Oncology Clinical Practice Guidelines into B2, B3 versus A, AB, B1 and B3 versus A, AB, B1, B2, the level of agreement decreased slightly ( $\kappa = 0.5506$  and  $\kappa = 0.4929$ , respectively). Difficulties arose in distinguishing thymoma from thymic carcinoma. Within the thymoma subgroup, difficulties in distinction were seen within the B group.

**Conclusions:** Agreement in diagnosing thymic lesions is substantial when they are assessed by pathologists with experience of these rare tumours. Digital pathology decreases the turnaround time and facilitates access to what is essentially a multinational resource. This platform provides a template for dealing with rare tumours for which expertise is sparse.

**Keywords:** interobserver variation, thymoma, tumour classification, whole slide imaging

Address for correspondence: MA den Bakker, Department of Pathology, Maasstad Hospital, PO Box 9100, 3007 AC Rotterdam, The Netherlands. e-mail: bakkerma@maasstadziekenhuis.nl

© 2020 The Authors. *Histopathology* published by John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## Introduction

Thymic epithelial tumours (TETs) are rare in routine practice. The estimated incidence of TETs in The Netherlands is 3.2/100 000.<sup>1</sup> These tumours are categorised according to the 2015 World Health Organization (WHO) classification into thymoma and thymic carcinoma. Thymomas are further divided into five major subtypes (A, AB, B1, B2, and B3) and rarer other thymomas. Segregation into categories is based on the relative proportions of the non-neoplastic lymphocytes, the proportions and cytological features (degree of atypia) of the neoplastic epithelial cells, and the resemblance of the tumour to the normal thymic architecture.<sup>2–4</sup> Although the WHO classification provides a number of well-defined categories, the diagnosis remains challenging, owing to both the rarity and the diversity of the tumours.

Several interobserver variability studies have evaluated the reproducibility of the WHO classification, often in combination with other classification systems, such as the Bernatz classification<sup>5</sup> and the Suster and Moran<sup>6</sup> system.<sup>7</sup> The level of interobserver agreement varies widely within the studies, with  $\kappa$ -values ranging between 0.37 and 0.87. Most difficulties were encountered in the subclassification of type B thymomas.<sup>8–12</sup> In recognition of these classification difficulties, diagnostic panels have been set up to improve consistency and aid the harmonisation of reporting.

In The Netherlands, a thymoma panel was initiated in 2009. The panel acts as a 'virtual panel' of 13 pathologists with a special interest in thymic pathology. As this is a virtual panel, its members do not physically meet but evaluate digitised (scanned) slides [whole slide imaging (WSI)] by the use of virtual microscopy. Most panel members are from The Netherlands, but there are also experts from the USA, Germany, and the UK. The panel provides subtyping of the tumour according to the WHO classification within an anticipated turnaround time of 2 weeks. To this end, the diagnoses of the panel members are tabulated and, if at least seven panel members have formulated an opinion, a consensus diagnosis based on the majority diagnosis is established and reported to the submitter. Panel members receive an alert when a case is finalised, and may review their diagnosis. In The Netherlands, submission of cases is voluntary, and it is the decision of the primary diagnosing hospital to request an opinion of the panel. The panel functions as a Dutch national reference panel for primary TETs and their differential diagnoses, and has been used as an example for panels in other areas of diagnostic pathology. Tumours

other than TETs, such as malignant lymphomas, germ cell tumours, and stromal thymic tumours, are also included.

The purpose of this study was to investigate the reproducibility of the WHO classification of TETs among a large group of international pathologists with interest and expertise in thymic pathology, assessed in a large series of successive cases submitted to the Dutch thymoma panel from 2011 until 2018 by the use of WSI.

## Materials and methods

Slides from submitted cases are digitally scanned at a central facility (Department of Pathology, Erasmus MC) and made available to panel members for external electronic review. The diagnoses are entered in a dedicated reporting system and summarised within a set time frame. Both haematoxylin and eosin (H&E)-stained sections and immunohistochemical stains from mediastinal masses from 49 hospitals in The Netherlands and Belgium (including seven university medical centres) were submitted to the panel between January 2011 and December 2018. Baseline demographic and clinical characteristics of the patients were provided by the submitting pathologist. There were no specific guidelines provided to the submitting institutions regarding the number of slides or additional stains required by the panel.

Tumour slides considered to be diagnostically relevant were scanned with the Hamamatsu NanoZoomer 2.0-HT (Hamamatsu Photonics K.K., Hamamatsu City, Japan). The scan magnification was set at  $\times 20$  or  $\times 40$  objective, which is equivalent to a resolution of approximately  $\times 200$  to  $\times 400$  with a normal microscope. The scans were then placed on a virtual platform named [www.pathpanel.org](http://www.pathpanel.org), where panel members had access to the virtual slides, which could be viewed in a suitable viewer such as NDP view 2 (Hamamatsu Photonics K.K.; see examples in Doc S1). Individual panel members entered their diagnosis in a categorical system according to WHO subtyping. After a 2-week evaluation period, a consensus diagnosis was formulated, which was reported back to the submitting pathologist. At the start of the panel, a consensus meeting was organised. In this meeting, potentially diagnostic criteria and practical issues were discussed, and adherence to WHO guidelines was strongly emphasised. For biopsies, a preferential diagnosis of 'thymoma not otherwise specified (NOS)' was advocated, and a subtype could be added in the comments. There was also the possibility of a second-round review

if additional stains were considered to be essential for a diagnosis at the time of the first review.

During the period of the study, 13 pathologists were members of the panel, although the composition of the panel did change owing to pathologists leaving the panel and new recruits. Cases were scored according to the WHO classification by the use of drop-down lists. Each panel member was thus able to render a preferred diagnosis individually without knowledge of the diagnoses of other members. In cases with varied histology, the proportions of different subtypes could be added in 10% increments (making a compulsory 100% sum). Alternatively, a 'non-thymic tumour' category could be used and specified in free text. Free-text comments could also be made. The criterion for formulating a consensus diagnosis was at least seven reviewers evaluating one case, and single outlying diagnoses were excluded from the final evaluation.

The data of all scoring pathologists were subjected to statistical analysis in Excel 2016 (v16.0) as part of Microsoft Office 2016 (Microsoft, Redmond, Washington, USA), and interobserver concordance (Fleiss kappa) was calculated.<sup>13</sup> (For calculations, see Doc S1.) The  $\kappa$ -value was used to calculate the reliability of agreement between the different pathologists when assigning categorical ratings to a number of given cases. A  $\kappa$ -value of  $\leq 0.20$  was regarded as poor agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as almost perfect agreement.<sup>14</sup> To gain further insights, clinically relevant diagnostic categories (A, AB, B1 versus B2, B3, and A, AB, B1, B2 versus B3) were defined, and  $\kappa$ -values between these groups were also calculated. Within the subgroups, the presence of a type B2 and/or type B3 minor component was considered in the calculations. All cases available in the panel were used for statistical analysis if at least seven pathologists participated in the scoring.

For descriptive purposes, the overall agreement was divided into five different categories: total agreement ranging from 95% to 100%, majority ranging from 75% to 94% agreement, consensus ranging from 50% to 74% agreement, trend ranging from 25% to 49% agreement, and lack with <25% agreement.

To illustrate the differences between previous biopsy and resection specimen, (consensus) diagnoses for both were listed separately and compared with each other.

## Results

Over a period of 8 years, 305 tumours, consisting of 90 biopsies and 215 resection specimens, were

**Table 1.** Baseline demographic and clinical characteristics

Characteristic	N = 305
Sex, n (%)	
Female	140 (45.9)
Male	165 (54.1)
Age at diagnosis (years), median (range)	61 (16–88)
Procedure, n (%)	
Biopsy	90 (29.5)
Resection specimen	215 (70.5)
Number of pathologists diagnosing a case: mode (range)	9 (7–12)

reviewed with a panel-based virtual microscopy approach. The baseline data of the submitted cases are summarised in Table 1. The predominant diagnosis for submitted cases was thymoma (70%, with type AB being the most prevalent diagnosis), followed by thymic carcinoma and benign thymic lesions, such as thymic cysts (Table 2). In most cases, the panel

**Table 2.** Distribution of diagnoses in thymoma panel submissions

WHO type	No.	Percentage
A	36	11.8
AB	80	26.2
B1	34	11.1
B2	41	13
B3	27	8.9
MNT-LS	18	5.9
Other thymoma	3	1
Thymic carcinoma	33	10.8
Carcinoid	1	0.3
Germ cell tumour	2	0.7
Lymphoma	1	0.3
Metastasis	7	2.3
Benign lesion	14	4.6
No consensus diagnosis	8	2.6

MNT-LS, micronodular thymoma with lymphoid stroma; WHO, World Health Organization.

**Table 3.** Thymoma panel assessment, overall agreement

Agreement, <i>n</i> (%)	Majority, <i>n</i> (%)	Consensus, <i>n</i> (%)	Trend, <i>n</i> (%)	Lack, <i>n</i> (%)
62 (20.3)	96 (31.5)	103 (33.8)	43 (14.1)	1 (0.3)

members gave no more than two different diagnoses for a single case (range, 1–7). The modal number of diagnosing pathologists for a case was nine (range, 7–12). A summary of the level of overall agreement is given in Table 3. The overall level of agreement was substantial ( $\kappa = 0.6762$ ), and differed between resection specimens ( $\kappa = 0.7281$ ) and biopsies ( $\kappa = 0.5955$ ). When thymomas were subclassified according to the European Society for Medical Oncology (ESMO) clinical practice guidelines for diagnosis, treatment and follow-up<sup>15</sup> into groups including B2, B3 versus A, AB, B1 and B3 versus A, AB, B1, B2, the level of agreement decreased from substantial to moderate ( $\kappa = 0.5506$  and  $\kappa = 0.4929$ , respectively) (Table 4).

The highest level of agreement was reached for thymoma types A and AB for biopsies and resection specimens ( $n = 71$ ;  $\geq 80\%$  per category). Examples with perfect agreement are shown in Figure 1. Differences in diagnoses arose in distinguishing thymoma ( $n = 8$  for B3 and  $n = 5$  for A;  $\geq 10\%$  per category)

from thymic carcinoma, and in distinguishing thymic carcinoma from metastasis ( $n = 8$ ;  $\geq 10\%$  per category). Within the thymoma subgroups, variation in subtyping was especially prevalent within the B group ( $n = 32$  for B1 versus B2 and  $n = 17$  for B2 versus B3;  $\geq 10\%$  per category).

In a further step, the consensus diagnosis of the biopsies was compared with the consensus diagnosis of the available matching resection specimen. In order to prevent bias, panel members were not aware of cases that had been previously biopsied and had been submitted to the panel. Of the 90 available biopsies, 15 cases could be linked to resection specimens that were submitted to the panel. There was 73% agreement between biopsy and resection in a predominant pattern, increasing to 87% (A, AB, B1 versus B2, B3) and 100% (A, AB, B1, B2 versus B3) for groupings in the ESMO treatment guidelines (Table 5).

## Discussion

Our study demonstrates substantial interobserver correlation for typing TETs when they were assessed by an international group of pathologists with experience in dealing with these rare tumours. However, because of the rarity of thymic epithelial tumours

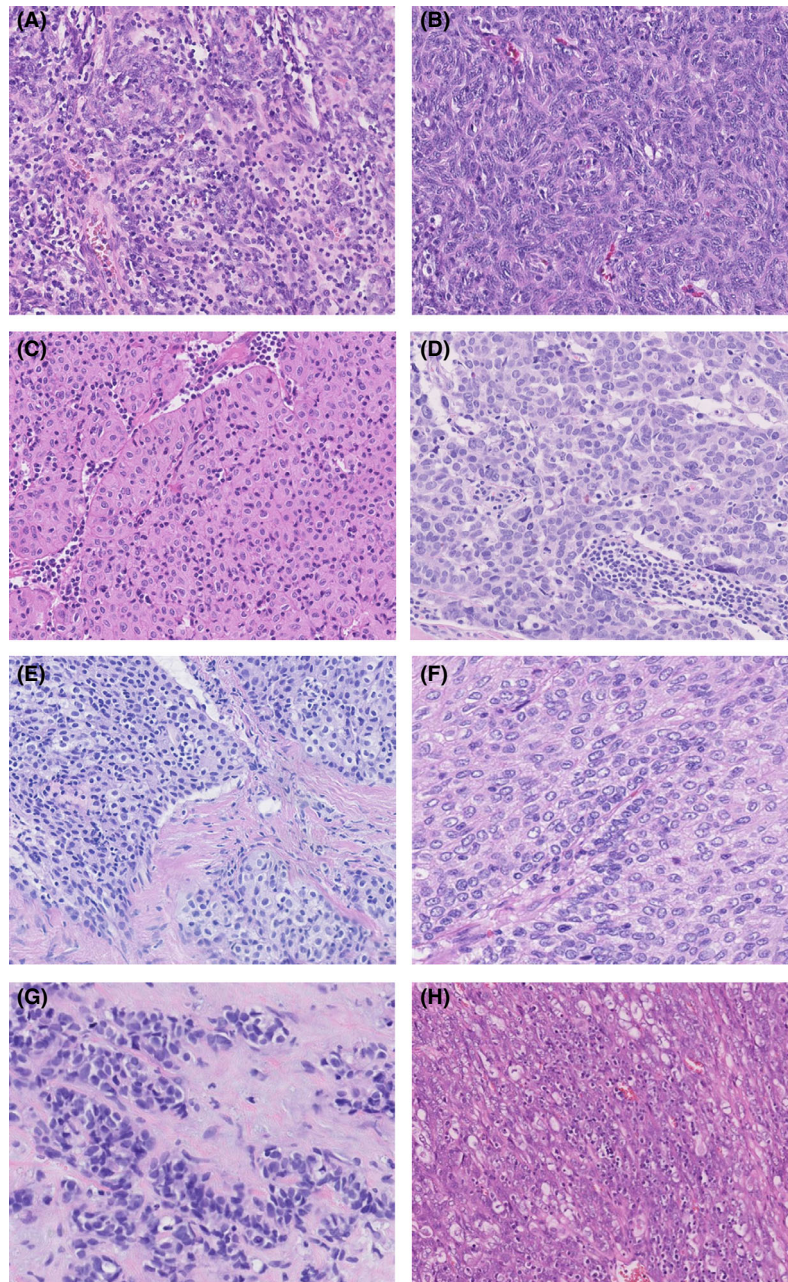
**Table 4.** Kappa values calculated for all specimens, and separately for biopsies and resection specimens

Diagnoses	Percentage agreement ( $p_a$ )	Percentage chance agreement ( $p_c$ )	$\kappa$ (coefficient) (95% CI)
All diagnoses			
Combined	0.8855	0.6466	0.6762 (0.6416–0.7108)
Biopsy	0.7945	0.492	0.5955 (0.5381–0.6530)
Resection	0.9224	0.7147	0.7281 (0.6615–0.7946)
Thymoma split 1			
Combined	0.7877	0.5275	0.5506 (0.5134–0.5879)
Biopsy	0.682	0.4919	0.374 (0.2933–0.4547)
Resection	0.824	0.5413	0.6163 (0.5754–0.6573)
Thymoma split 2			
Combined	0.7877	0.6951	0.4929 (0.4352–0.5506)
Biopsy	0.7749	0.6578	0.3423 (0.2348–0.4497)
Resection	0.8695	0.7092	0.5512 (0.4940–0.6083)

CI, confidence interval.

Kappa values were calculated for thymoma subgroups, with thymomas divided into two categories with different therapeutic consequences. Split 1 refers to (A, AB, B1) versus (B2, B3), and split 2 refers to (A, AB, B1, B2) versus (B3).





**Figure 1.** Examples of thymomas from the thymoma panel; cases A–D were scored with high consensus rates, and for cases E–H there were split opinions. A, Encapsulated mediastinal mass scored by seven pathologists, with 100% consensus for type AB thymoma. B, Resected mediastinal tumour scored by 10 pathologists, with 100% consensus for type A thymoma. C, Resected thymic tumour scored by seven pathologists, with 99% consensus for type B3 thymoma. (1% type B2). D, Thymic resection unanimously scored by seven pathologists as thymic carcinoma. E, Biopsy specimen with a thymic epithelial tumour weakly staining for p40. CD117 and terminal deoxynucleotidyl transferase were negative. The specimen was scored by nine pathologists, with an outcome of 55% type B3 thymoma and 45% thymic carcinoma. F, Resected encapsulated mediastinal tumour scored by nine pathologists, with an outcome of 74.7% type B3 thymoma, 22.2% thymic carcinoma, and 3.33% type B2 thymoma. G, Resected anterior mediastinal tumour. The tumour was positive for p40, cytokeratin (CK) 5, CK19, Pax8, and CD117, and negative for CK7, thyroid transcription factor-1, napsin A, chromogranin, and synaptophysin. It was scored by nine pathologists, with an outcome of 77.78% thymic carcinoma and 22.22% type A thymoma. H, Resected multilobular tumour from the anterior mediastinum. The tumour was partly positive for CD5 and CD99, and negative for CD117. It was scored by 10 pathologists, with an outcome of 58% type B3 thymoma, 23% thymic carcinoma, 10% type AB thymoma, 8% type B2 thymoma, and 1% type A thymoma.

**Table 5.** Diagnoses in matched biopsy–resection specimens

	No. of assessors	Biopsy	No. of assessors	RResection
Case 1	9	<b>65% MNT-LS</b> (25% AB, 10% A)	10	<b>72.22% MNT-LS</b> (16.66% A, 11.11% AB)
Case 2	7	<b>92.85% B1</b> (7.15% B2)	7	<b>51.43% B1</b> (22.86% AB, 25.71% B2)
Case 3	7	<b>47.77% B2</b> (30% B1, 22.22% AB)	9	<b>100% AB</b>
Case 4	5	<b>80% B1</b> (20% B2)	9	<b>64% B2</b> (36% B1)
Case 5	8	<b>84.29% Thymic carcinoma</b> (15.71% B3)	7	<b>88.75% Thymic carcinoma</b> (11.25% B3)
Case 6	8	<b>76.66% B1</b> (22.22% tNOS, 1.11% NTT)	9	<b>66.25% B1</b> (33.75% B2)
Case 7	8	<b>78.5% A</b> (21.5% AB)	7	<b>100% A</b>
Case 8	8	<b>66.66% B1</b> (22.22% NOS, 11.11% B2)	9	<b>36.25% B1</b> (25% AB, 20% B2, 12.5% other, 6.25% B3)
Case 9	10	<b>56.25% B2</b> (18.75% B3, 12.5% tNOS, 12.5% AB)	8	<b>88% B2</b> (12% B1)
Case 10	8	<b>88.88% A</b> (11.11% tNOS)	9	<b>92.5% A</b> (7.5% B3)
Case 11	8	<b>68.18% MNT-LS</b> (22.72% A, 9.09% normal thymus)	11	<b>100% MNT-LS</b>
Case 12	8	<b>68.75% B1</b> (25% tNOS, 6.25% B2)	8	<b>66.25% B2</b> (17.5% B1, 12.5% AB, 3.75% B3)
Case 13	10	<b>94.44% A</b> (5.56% AB)	9	<b>100% A</b>
Case 14	8	<b>95% A</b> (5% AB)	10	<b>43.75% A</b> (12.5% AB versus NTT versus other, 10% B3, 7.5% tNOS, 1.25% B2)
Case 15	6	<b>40% MNT-LS versus tNOS</b> (20% AB)	5	<b>65% AB</b> (30% MNT-LS, 5% B1)

MNT-LS, multinodular thymoma with lymphoid stroma; NTT, non-thymic tumour; tNOS, thymoma not otherwise specified. Consensus diagnoses are given in bold.

and the spectrum of thymoma subtypes, the diagnosis remains a challenge for many, even experienced, pathologists, as demonstrated by different interobserver variability studies. In these studies, the range of agreement ranged from fair to substantial to almost perfect, with  $\kappa$ -values between 0.34 and 0.87. This might be explained by the number of evaluating pathologists, the heterogeneity of the thymic lesions, the experience of the assessors, and the amount and variability of cases and additional stains evaluated.<sup>8–</sup>

<sup>10</sup> Evaluation of TETs by a small number of assessors will usually lead to less interobserver variability, particularly if the assessors are within a single institute and frequently discuss cases.<sup>11,16,17</sup>

To maximise accuracy in diagnosing thymic tumours, changes were introduced in the 2015 WHO classification,<sup>3</sup> as exemplified in detail in the assessment of the International Thymic Malignancy Interest Group statement on the WHO histological classification.<sup>3</sup> Revision and refinement of histological and immunohistochemical diagnostic criteria should lead to more robust and reproducible subtyping of

thymomas and distinction between thymomas and thymic carcinomas. To these ends, major and minor diagnostic criteria were introduced, as were recommendations on the use of immunohistochemical markers for the diagnosis of thymomas with ambiguous histology and thymic carcinomas.

The good level of agreement in our study might be explained by the high level of expertise of pathologists from different countries in Europe and the USA, and the strict adherence to diagnostic criteria, indicating that the current WHO criteria for classification of TETs are sufficiently reproducible for diagnostic purposes. However, even with this level of agreement, cases remain that are challenging to classify. In the overall group of biopsies and the resection specimens, the main difficulties arose in distinguishing between thymic carcinoma and metastasis, between type B3 thymoma and thymic carcinoma and between B3 thymoma and type A thymoma. This might be explained by the degree of histological overlap. Although obvious high-grade thymic carcinomas are readily diagnosed, those cases that show organotypic

features, such as lobulation and septation, or those that have limited cytonuclear atypia, overlap considerably with type B3 carcinoma. Furthermore, insufficient sampling, heterogeneity of the tumour or lack of additional stains to differentiate between thymoma, thymic carcinoma and metastasis can make the diagnostic process more challenging, as reported by some pathologists. When thymoma subgroups were evaluated, lower  $\kappa$ -values were seen, with most difficulties being encountered in differentiating between different type B (B1, B2, and B3) thymomas. These findings were also reported by Rieker *et al.*<sup>11</sup> Within this subgroup weighted  $\kappa$ -values declined from 0.87 to 0.49.

Kappa values differed between resection specimens and biopsies, with lower values for biopsies. This might be explained by the limited amount of available tissue in a biopsy, and the fact that thymic tumours can be very heterogeneous. This can make it difficult to provide a definitive classification on a small core biopsy, and caution is advocated in this situation.<sup>4</sup>

However, panel responses vary in these situations, with some panel members rendering a diagnosis of thymoma NOS on biopsy specimens, and others subtyping if feasible. To evaluate this recommendation, we linked the biopsies to the available resection specimens. Although the consensus diagnoses of the biopsies corresponded with the findings of the resection specimens in most matched cases (11 of 15 cases; 73%), we noticed differences in subtyping of thymoma subgroups. This might be explained by sampling bias, and the presence of different thymic subtypes within one tumour, and is especially important to be aware of in cases in which biopsy subtyping influences clinical management.

As some therapeutic decisions are based on groupings of thymoma subtypes, we performed a subclassification of thymomas according to the ESMO clinical practice guidelines for diagnosis, treatment and follow-up into groups including (B2, B3) versus (A, AB, B1) and (B3) versus (A, AB, B1, B2). As stated in the guidelines, Masaoka–Koga stage IIA resected thymomas (A, AB, B1, and B2) will be followed up without further treatment, whereas postoperative radiotherapy is recommended for resected type B3 thymomas. In stage IIB, the recommendation for postoperative radiotherapy is extended to also include resected type B2 thymomas.<sup>15</sup> Therefore, in conjunction with clinical stage, thymoma subtyping impacts on therapeutic decision-making. When thymomas were clustered into (B2, B3) versus (A, AB, B1) and (B3) versus (A, AB, B1, B2), the level of agreement decreased to moderate ( $\kappa = 0.5275$  and  $\kappa = 0.4929$  respectively), further emphasising the difficulty in segregating the B subgroups.

A limitation of the study is that, for most cases, the Dutch thymoma panel received H&E-stained sections and a small subset of immunohistochemical slides only, and rarely received tissue blocks or unstained glass slides that could be used for additional stains. However, a (consensus) diagnosis was established with the available material in most cases. Furthermore, in cases deemed to have insufficient material for a confident diagnosis, the panel could request additional stains to establish a diagnosis. Therefore, despite these limitations, the interobserver correlation was good, but could be enhanced further by expanding and standardising the panel of immunohistochemical stains, especially for subtyping small needle core biopsies and complex resection specimens.

Future molecular studies might aid in the distinction between different thymoma subtypes and between thymomas and thymic carcinomas. In this respect, alterations of *GTF2I*, which has a high mutation frequency in type A and AB thymomas when compared to other thymoma subtypes and thymic carcinomas, is a good example.<sup>18,19</sup> Other examples are mutations in *KIT* or *TP53* in thymic carcinomas.<sup>20,21</sup>

In recent years, the application of digital pathology using scanned slides (WSI) has increased considerably, and has been shown to be a reliable alternative to conventional microscopy.<sup>22</sup> Development and improvement of scanning technology has resulted in greatly reduced scanning times, increased resolution, and the introduction of user-friendly interfaces. A major advantage of virtual slide technology is easy and convenient global sharing of rare tumour cases such as TETs, as demonstrated by the Dutch thymoma panel, which, in The Netherlands, was the first of its kind. A similar study utilising WSI for evaluating TETs was reported previously by Wang *et al.*<sup>9</sup> The reported overall  $\kappa$ -value was lower, at 0.39 compared with 0.67 in our study. This is probably due to differences in the number and selection of cases, and in the number of evaluating pathologists. It is of note that in the study of Wang *et al.* observer agreement improved in a second round after discussion of cases, which means that interpretation of criteria may have originally differed between pathologists, and that strict adherence is required for accurate diagnosis of TETs.

Submitting material from rare tumours to an expert panel can increase the quality of the diagnosis. With the addition of tissue blocks or unstained sections, additional ancillary studies can be performed in the coordinating institution and shared in an online platform by the use of WSI. This allows experts around the world to gain access to these rare lesions and generate the most accurate diagnosis.



Furthermore, making WSI of all cases with diagnoses available as an online resource could also provide an excellent learning tool.

In conclusion, this study shows that, with adherence to the WHO guidelines, substantial agreement on the classification of TETs is achieved. This study further indicates that the current WHO criteria allow good categorisation of TETs. In addition, further confirmation of the diagnostic accuracy and rapid turnaround time of WSI is provided.

## Conflicts of interest

The Author(s) declare(s) that there is no conflict of interest.

## Funding

No funding was received for this study.

## Author contributions

M. A. den Bakker, J. von der Thüsen, and J. L. Wolf: conceptualisation and methodology. J. L. Wolf: data curation, formal analysis, original draft preparation, and visualisation. M. A. den Bakker, J. von der Thüsen, and J. L. Wolf: investigation. M. A. den Bakker and J. von der Thüsen: supervision. J. L. Wolf, F. van Nederveen, H. Blaauwgeers, A. Marx, A. G. Nicholson, A. C. Roden, P. Ströbel, W. Timens, A. Weissferdt, J. von der Thüsen, and M. A. den Bakker: writing, reviewing, and editing.

## References

- de Jong WK, Blaauwgeers JL, Schaapveld M, Timens W, Klinkenberg TJ, Groen HJ. Thymic epithelial tumours: a population-based study of the incidence, diagnostic procedures and therapy. *Eur. J. Cancer* 2008; **44**: 123–130.
- Marx A, Chan JK, Coindre JM *et al.* The 2015 World Health Organization Classification of Tumors of the Thymus: continuity and changes. *J. Thorac. Oncol.* 2015; **10**: 1383–1395.
- Marx A, Ströbel P, Badve SS *et al.* ITMIG Consensus Statement on the Use of the WHO Histological Classification of Thymoma and Thymic Carcinoma: refined definitions, histological criteria and reporting. *J. Thorac. Oncol.* 2014; **9**: 596–611.
- Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG, eds. *World Health Organization classification of tumours of the lung, pleura, thymus and heart*. 4th ed. Lyon: IARC Press, 2015.
- Bernatz PE, Harrison EG, Clagett OT. Thymoma: a clinicopathologic study. *J. Thorac. Cardiovasc. Surg.* 1961; **42**: 424–444.
- Suster S, Moran CA. Thymoma, atypical thymoma, and thymic carcinoma. A novel conceptual approach to the classification of thymic epithelial neoplasms. *Am. J. Clin. Pathol.* 1999; **111**: 826–833.
- Roden AC, Yi ES, Jenkins SM *et al.* Reproducibility of 3 histologic classifications and 3 staging systems for thymic epithelial neoplasms and its effect on prognosis. *Am. J. Surg. Pathol.* 2015; **39**: 427–441.
- Oselin K, Girard N, Lepik K *et al.* Pathological discrepancies in the diagnosis of thymic epithelial tumors: the Tallinn-Lyon experience. *J. Thorac. Dis.* 2019; **11**: 456–464.
- Wang H, Sima CS, Beasley MB *et al.* Classification of thymic epithelial neoplasms is still a challenge to thoracic pathologists: a reproducibility study using digital microscopy. *Arch. Pathol. Lab. Med.* 2014; **138**: 658–663.
- Verghese ET, den Bakker MA, Campbell A *et al.* Interobserver variation in the classification of thymic tumours—a multicentre study using the WHO classification system. *Histopathology* 2008; **53**: 218–223.
- Rieker RJ, Hoegel J, Morresi-Hauf A *et al.* Histologic classification of thymic epithelial tumors: comparison of established classification schemes. *Int. J. Cancer* 2002; **98**: 900–906.
- Zucali PA, Di Tommaso L, Petrini I *et al.* Reproducibility of the WHO classification of thymomas: practical implications. *Lung Cancer* 2013; **79**: 236–241.
- Fleiss J. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 1971; **76**: 378–382.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
- Girard N, Ruffini E, Marx A, Faivre-Finn C, Peters S, ESMO Guidelines Committee. Thymic epithelial tumours: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 2015; **26**(Suppl. 5): v40–v55.
- Chen G, Marx A, Chen WH *et al.* New WHO histologic classification predicts prognosis of thymic epithelial tumors: a clinicopathologic study of 200 thymoma cases from China. *Cancer* 2002; **95**: 420–429.
- Park MS, Chung KY, Kim KD *et al.* Prognosis of thymic epithelial tumors according to the new World Health Organization histologic classification. *Ann. Thorac. Surg.* 2004; **78**: 992–997; discussion 997–998.
- Okuda K, Moriyama S, Haneda H, Kawano O, Nakanishi R. Specific mutations in thymic epithelial tumors. *Mediastinum* 2017; **1**: 16.
- Radovich M, Pickering CR, Felau I *et al.* The integrated genomic landscape of thymic epithelial tumors. *Cancer Cell* 2018; **33**: 244–258.
- Enkner F, Pichlhofer B, Zaharie AT *et al.* Molecular profiling of thymoma and thymic carcinoma: genetic differences and potential novel therapeutic targets. *Pathol. Oncol. Res.* 2017; **23**: 551–564.
- Roden AC. Molecular changes are infrequent in thymic carcinomas but might represent targetable mutations. *Mediastinum* 2017; **1**.
- Mukhopadhyay S, Feldman MD, Abels E *et al.* Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *Am. J. Surg. Pathol.* 2018; **42**: 39–52.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Doc S1.** Calculation method and examples of the panel platform.